Lustre WAN over 100Gbps

Abhinav Thota, Robert Henschel, Thomas William,
Christian Terboven, Hans-Juergen Schnitzer, Jennifer
Schopf

Nov 18, 2013

RESEARCH TECHNOLOGIES
INDIANA UNIVERSITY
University Information Technology Services

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

Why is Lustre WAN over 100G interesting

- Lustre is a distributed parallel file system
- Highly scalable
- Popular choice among HPC users
- DC-WAN is one of the Lustre file systems at IU, mounted at many locations across the world at various points in time (NCSA, Pittsburgh, Mississippi State, Tucson, TACC, Dresden, Aachen, etc.)
  - Lets researchers access remote data as if that data were stored locally
- Multiple organizations can work on the same data from multiple locations
- Acceptable performance – depending on the application I/O characteristics
- Convenience – data stored in an accessible central location
- Only one copy of the data
- Unattractive alternatives
  - Copy/move data with scp, Globus/GridFTP or similar
  - Version control
  - Fedex

RESEARCH TECHNOLOGIES
INDIANA UNIVERSITY
University Information Technology Services

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

Past experience with Lustre over wide area

- We have been doing this for a long time

- 10G between IU and ORNL in '06

- 100G between Dresden and Freiberg in '10

Dresden Freiberg 🛣 100 Gbit/s

- 100G between Indianapolis and Seattle in '11

So far… on non-dedicated transatlantic links

- Started with the BLASTN application
  - Popular application among genomics/biology/bioinformatics groups
  - Large common input data sets
  - Cache effect when reading from and writing to lustre file system
- Mounted DC-WAN at Blacklight, Aachen and Dresden
- On each of these machines: ran BLASTN R+W from/to DC-WAN and local lustre or NFS file systems
- On the next slide, ratio of runtimes DC-WAN vs. local vs. SCP on each system
  - scp in input data and scp out output data
- Observe little difference between the DC-WAN and SCP time to completions
  - DC-WAN wins as it is more convenient
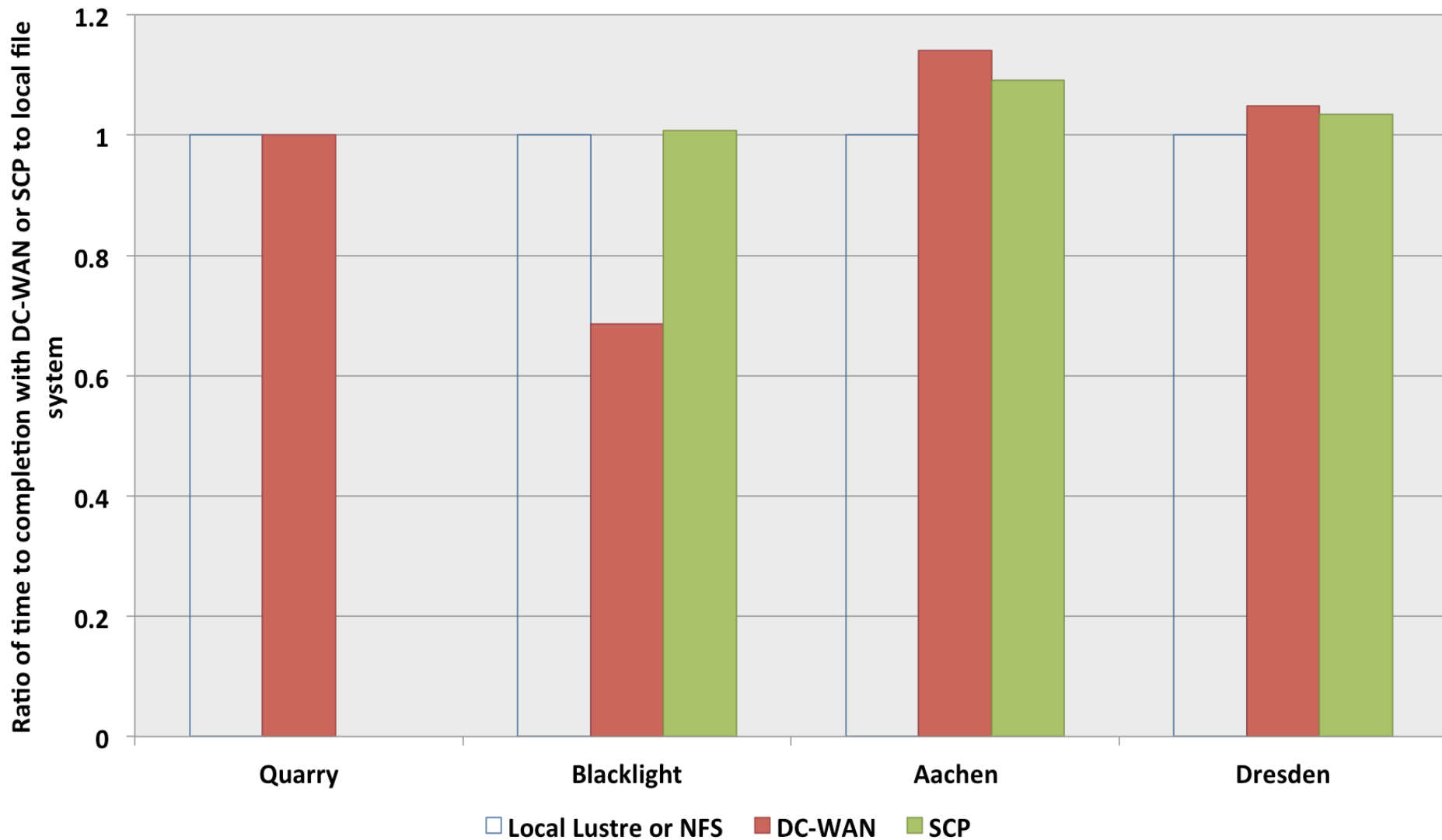- Fluctuating bandwidth and latency without a dedicated link

Ratio of time to completion of BLASTN with DC-WAN and SCP to local lustre or NFS file system

In the pipeline… on the dedicated 100G link

- Experiment with a list of applications
  - BLAST
  - Data Analysis using Virtual Reality
  - Combustion Research
  - AMS Data Processing
  - Comparing UNICORE FTP and Lustre WAN
  - Analyzing Microscopy Data
  - Distributing NCGAS Workflows
- Get a thorough understanding of the issues involved
- Compare the Lustre workflow to the alternatives
- Make a recommendation based on application characteristics

RESEARCH TECHNOLOGIES
INDIANA UNIVERSITY
University Information Technology Services

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY